Analysis on the Risk and the Categorization on Test Automation in Sri Lankan Software Industry

Sakthi Sundaralingam
(Reg. No.: MS12907458)
M.Sc. in Information Technology

Supervisor: Mr. U. U. Samantha Rajapaksha

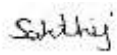
May 2021

Department of Information Technology
Faculty of Computing
Sri Lanka Institute of Information Technology

Declaration

This thesis is presentation of my original work. This submission in my own work and has been generated by me as the result of my own work. I hereby declare that to my best knowledge not included any published or others material.

Certified by



Signature:……..

Date: 21-August-2021

The above candidate has carried out research for the M.Sc. in IT thesis under my supervision.

Name of Supervisor: Mr. U. U. Samantha Rajapaksha

Signature:……..

Date:……..

Abstract

Delivering quality software to customer is the key objective of software industry. One of the essential fragment of life cycle of software is software testing. In software testing test automation is playing a major role. Test automation is an art which needs to be managed and executed properly in order to provide quality software. If test automation cannot be practiced in proper way the delivery of the software quality would impact directly and leads to loss of customer, which is a failure of business. Test automation contain several phases from scoping to maintenance. Each phase has several steps which require right analyzing, decision making and operating.

Test automation has several problems which needs to address in each stage. Test automation cause several issues when execute test automation in a company. All these issues need to be handled by different people, therefore initially issues need to be identified and classified and then solve properly.

This research is to identify the improvements to categorize the problems automatically and find the solution for the problem in test automation process and hence to practice the test automation in healthier way in order to achieve better software quality. Test automation issue are analyzed and the solutions are proposed. On which stage, the test automation is causing problems and how to solve them are recommend in this research, Test automation issues are categorized and under relevant category therefore issues can be solved speedily.

The automation cause huge amount of issues. Test automation is running over nigh in most of the companies and it result lots of issues. These issues can't be analyzed one by one and allocated to relevant people. The implementation of this research will categorize the issues under relevant category and which would lead for speedy allocation under relevant person and solution. This research produced a system to categorize the test automation problems in automatic way using text classification approach. The issues are passed as sentence and they are categorized under the relevant category to fix them quickly. The sentences are preprocessed and conducted feature selection using filter methods and predict under appropriate category. The issue has been cleaned in preprocess stage. Implemented LSTM base algorithm using filter method to categorize the issues.

Find the solution for the problem in test automation process and how to practice the test automation in healthier and proper way in order to achieve better software quality. In this research an implementation to categorize test automation problems were formed. Recommendation and solutions are proposed on test automation which would aid to practice test automation in better way and that would leads to better software quality delivery.

Acknowledgment

I would like to thank Mr. U. U. Samantha Rajapaksha of Sri Lanka Institute of Information Technology for the guidance and the extended support provide for this research. I also want to thank to each and every one who participate in the review and survey.

Table of Contents

Declaration i
Abstract ii
Acknowledgment iv
Table of Contents 1
List of Figures 3
List of Tables 4
Abbreviation 5
Chapter 1 Introduction 1
1.1 Background 1
1.2 Problem definition 2
Categorizing test automation problem under relevant category 2
Bottleneck in decision making, planning, designing, analysis, reporting, measuring and maintenance of test automation. 3
1.3 AIM 5
1.4 Research Objective 5
1.4.1 Main objective 5
1.4.2 Sub objectives 5
1.5 Research question 6
Chapter 2 Literature Review 7
2.1 Test automation practices 7
2.2 Test automation risk and classification 12
2.3 Software quality 13
Chapter 3 Methodology 14
3.1 Process overview 14
3.2 Data collection 15
3.3 Population 16
3.4 Sample 16
3.5 Research framework 17
3.6 Test automation process 18
3.7 Classification categories 20
3.8 Text classification approach 20
3.9 Why LSTM 22
3.10 Solution system diagram 23
3.11 Implementation user interface 23
3.12 Preprocessing data 24
3.13 Normalization 25

3.14 Tokenization 25
3.15 Vectorization 25
3.16 Tokenizer pickle 25
3.17 Optimization 25
3.18 Feature selection 25
Chapter 4 Result and discussion 26
4.1 Reliability analysis 26
4.2 Variables 26
4.3 Correlation matrix 28
4.4 Demographic result 30
4.5 Quantitative result 32
4.6 Quantitative result methodology wise 36
4.7 Qualitative result on risks in test automation 37
4.8 Qualitative result on solution and recommendations to test automation problems and risks 40
4.9 Result of solution 42
4.10 Accuracy graph 43
4.11 Loss graph 44
4.1 Recall graph 44
4.2 Precision graph 45
4.3 Evaluation 45
4.3.1 Accuracy evaluation 46
4.3.2 Loss evaluation 46
4.3.3 Classification report 47
4.3.4 Confusion matrix 47
4.3.5 Test cases 47
Chapter 5 Conclusion and future work 50
5.1 Conclusion 50
5.2 Future work 50
References 51
Glossaries 53
Questionnaire and received result 54
Code 56

List of Figures

Figure 3.1 Research methodology 14
Figure 3.2: Data collection.....16
Figure 3.3: Research framework.....17
Figure 3.4: Classifcation 20
Figure 3.5: System diagram 23
Figure 3.6: Interface.....23
Figure 3.7: Processing Data 24
Figure 4.1: Stream working 30
Figure 4.2: Years of experience 31
Figure 4.3: Level of work 31
Figure 4.4: Methodology mostly worked..... **Error! Bookmark not defined.**32
Figure 4.5: Risk analysis and delivery 33
Figure 4.6: Scope, plan and requirement 34
Figure 4.7: Operations 35
Figure 4.8: Maintenance, review and management 35
Figure 4.9: Methodology vs Risk analysis and delivery 36
Figure 4.10: Methodology vs Scoping, planning & requirement.....36
Figure 4.11: Methodology vs Operations.....37
Figure 4.12: Methodology vs Maintenance, review and management.....37
Figure 4.13: Solution accuracy 43
Figure 4.13: Solution loss 44
Figure 4.14: Solution recall.....44
Figure 4.15: Solution precision.....45
Figure 4.16: Confusion matrix.....47

List of Tables

Table 4.1: Variables 26
Table 4.2: Correlation 28
Table 4.3: Accuracy Comparison.....46
Table 4.4: Loss Comparison 46
Table 4.5: Classification Report……...47

Abbreviation

LSTM: Long Short Term Memory

HTML: Hyper Text Markup Language

UI: User Interface

UI/UX: User Interface / User Experience

CI: Continuous Integration

CD: Continuous Delivery

ROI: Return on Investment

RNN: Recurrent Neural Network

TC: Test Case

BERT: Bidirectional Encoder Representations from Transformers

Tf-idf: Term Frequency-Inverse Document Frequency